

THE NORC LIBRARY AS INFORMATION CENTER AND DATA ARCHIVE

by
Patrick Bova
National Opinion Research Center

INTRODUCTION

The objective of this paper is to provide a description of the origin of the National Opinion Research Center (NORC) Library and its information base, the user community it serves and some of the problems encountered, especially with the data archive. Although this paper describes the the NORC Library, we will also refer to NORC itself, since the Library is, in many ways, the public arm of the organization. The NORC Library is that part of the agency through which information passes to the public.

The NORC Library can best be described as a special library, in the sense that word is used to classify libraries. It is active as a conventional library with a book and reference collection closely aligned to the current research interest of the intramural research programs. It is the NORC information office which supplies all manner of information about NORC, NORC studies, and survey research in general, to a wide variety of publics. It is the data archive for NORC data and for data generated with NORC participation. It is definitely not the local service data archive for the University of Chicago, although many people acquire data from NORC for research as well as information about data located elsewhere. The University maintains a separate data facility (within the Division of the Social Sciences) which houses data from ICPSR and, interestingly enough in the present context, the University copies of the General Social Survey. The present form of the NORC Library is a direct result of the recent history and growth of NORC and of the social sciences.

HISTORY OF NORC

The history of NORC includes changes in the size of NORC and in the number, complexity and subject matter of studies undertaken. NORC was founded in 1941 as a place where the then new technique of opinion sampling could be applied in the neutral setting of a university. From 1941 to about 1960 the center had a modest budget and a

small staff in the Chicago headquarters and in a New York City office. NORC could be characterized in that period as a research center housed in an old mansion on campus; the staff was small enough to meet every day in the dining room of that house for coffee and cake.

The notable studies conducted during that period were the series of national surveys on the conduct of foreign affairs conducted for the U.S. Department of State between 1945 and 1957 (more about these later), the 1947 North-Hatt occupational prestige study, the Shirley Star study of popular attitudes toward mental illness, and a series of health-related surveys of which some are still being repeated by the Center for Health Administration Studies (CHAS) at the University.

A library existed in those days to serve the professional staff in Chicago and New York. One important and useful activity was the maintenance of a question or item index, by subject, of the foreign affairs studies conducted in the period from 1945 to April, 1957. This index contains the full question text with percentage marginals, plus some information on sample type and size. It is still in use today, although we are not confident that it is any longer complete. The item index is quite useful in accessing these studies (now at Roper Center), which number about 125, and since the marginals are included, the file is a research tool in itself, for certain kinds of analysis.

Besides the question index, the library maintained files of materials used in studies and a small book and journal collection closely related to the subject matter of the studies being done (we still have a concise collection of important health-related survey research done in the 1950s). In addition, rather extensive files of poll and survey results from many agencies then producing them were maintained, including foreign sources. These files still exist and are used from time to time, although we have not made an attempt to keep them up to date due to cost and low user interest. We intend to keep the files, of course, since they are probably a unique resource for that time period.

It appears that the librarian, a part-time staffer, worked mainly on the informational needs of the staff; the poll and survey results files were there because at that time NORC was issuing press releases much in the fashion that Gallup does now as well as publishing the journal Opinion News, which covered results from studies by NORC and other organizations. Public contact by the library was probably minimal, and whatever public contact there was at NORC was often a function of the study director. Even in the early 1960s, when the present librarian began working for NORC, the level of public information activity was quite low.

The 1960s changed NORC's style of public representation, just as the period changed other aspects of academe. At NORC several things occurred which had consequences for the library -- the number and complexity of studies in progress increased and the topics proliferated well beyond health studies, so that the library collections expanded and diversified. The NORC professional staff grew and changed -- new staff members came to depend on the library for historical information; as public contact increased, requests and questions were increasingly referred to the library.

THE CONTEMPORARY RECORD OF NORC

The brief preceding discussion has largely to do with the internal research programs at NORC. In 1963 NORC founded its Survey Research Service (SRS) to formally provide research services, chiefly data collection (but including all phases of survey research), to the social science community. The Center had previously contracted to conduct studies for extramural sponsors (a good example is the Stouffer communism study in 1954 in conjunction with the Gallup organization), but SRS was actively promoted as such. The major characteristic of SRS that should be noted is that in most cases NORC's role was technical. -- we usually had little to do with the initial planning of these studies and little to do with analyzing data and reporting the results. This characteristic has implications for the library in its information and data archive activities, as we shall see.

Coincident with the founding of SRS were the start of the Great Society poverty programs which included built-in evaluation appropriations. SRS was available to

carry out some of the massive data collection programs needed and became quite busy at it.

The public image of the social sciences changed in the 1960s, too -- one could say that the social sciences wanted the public to be more attentive to its results, and the public apparently began to be, helped along by such magazines as TRANS-action and Psychology Today. There were other currents -- the boom in higher education and in the graduate student population, the increased sophistication in the use of the computer in the soft sciences -- that had effects on the NORC Library and forced it to be more responsive to a larger and larger audience.

One development in the 1960s had a great impact on the NORC Library: the data archive movement. Prior to 1960 people were certainly sharing data but in a more person to person as opposed to an agency to person manner. Times were, after all, simpler; there simply was not as much data or demand for them as we see now. What there was tended to be more straight forward, perhaps, than many studies are today. At NORC the first and probably most dramatic act related to data archiving was the depositing of the State Department studies in the Roper Public Opinion Research Center in 1959. This act is more than symbolic of our policy (if not our activity) on archiving at NORC which evolved during the 1960s: to deemphasize archiving. The act of releasing the State Department surveys was significant because it anticipated the revised policy.

Our decision to try, at least, to de-emphasize NORC as a data archive had to do with the fact that running an archive is an expensive proposition. Our data are NORC studies, many quite old and many in sad condition. The organization could not get funds to do the archiving properly, and could not provide the funds itself, so we decided to try to send those studies we felt were usable to a logical archive (we hoped to find appropriate archives which would house many of the studies).

Although data archiving has been deemphasized, we have not stopped archiving machine readable data. The old studies mentioned above -- although not all are worth equal attention -- still require much work before we would inflict them on any archive. We therefore provide sample survey data for secondary use, but we do not collect or keep data from other sources, ex-

cept those which may have been acquired for a specific project.

Just as our data archive activity is closely identified with NORC production, so is our information activity. And just as data are the result of a long line of activity, so too do our information services cover all aspects of these events, not only to NORC staff but also to the interested public.

The survey movement is enormously active these days, and more and more people are interested not only in the results of these many studies but also in the methods of conducting them. The Library responds to most requests for information on NORC projects or tries to get someone on the staff to respond. We also provide an introduction to NORC and survey research to school groups that might be interested in visiting NORC. The questions we receive cover the social sciences, but mainly stay within the purview of survey research. Questions have to do with methods (eg: question wording), procedures (eg: interviewer training) and with particular subject matter.

Our problem with using NORC generated information is the usual one -- access; at present our methods of retrieval are primitive, but we are working on solutions. The location of a study on a particular subject or method is just the start, depending on what information the user needs. In most cases, the results of a study are requested, and here we inform the user as best we can. If data are required, we supply them, assuming we have a copy. In any case, data are supplied only with sponsor or study director permission.

The data for extramural studies which reside at NORC are in an interesting limbo. Although most contracts call for public release of data after a set time (usually 2-3 years), our feeling is that contact with the original sponsor to seek permission to use the data serves useful purposes, not the least of which is to learn about the original analysis and to talk with someone who is presumably working in the same area. As time passes more and more of these data sets are "willed" to NORC. What usually happens is that the sponsor will get tired of giving permission and allows us blanket permission to make the data available to all comers, including archives. In some cases it turns out that we have the best, original copy of the data, since we will not have used the tape (except maybe to copy a few

times) and since many original sponsors lack the institutional setting needed to care for tapes and documentation for any length of time. These data sets are treated in the same manner as NORC's own data.

THE IMPACT OF USER DEMAND

We turn now to a consideration of how user demand affects the work of the Library (and perhaps how it should affect that work) and how the nature of survey research as a scientific enterprise may help us set some priorities. We will be concerned with two basic problems of documenting or archiving surveys at NORC: how to cope with the diversity of information requests which compete with and take precedence over the archiving and documenting of surveys, and how recent changes in data processing and analysis have created additional documentation problems.

One of the fundamental problems in the NORC Library is that we serve a wide variety of users with quite diverse information needs -- from high school students who need help with a debate topic, to quite sophisticated professionals who have serious and complex data requirements. As a result we quite often find ourselves reacting, in a passive way, to immediate demands. The careful work that is required to get a data set in shape for secondary use is quite often simply another thing to do in the long list of things to do, and most often just does not get done. It appears that in the hurly-burly of day to day routine, we take the pragmatic route only and try to satisfy those who are most insistent for service.

THE ROLE OF THE ARCHIVIST

Now, everyone deserves attention, one would presume the high schooler as much as the potential user of one of our surveys (which may lie "unarchived", in unusable shape). Why should we pay special attention to the archive function when limited resources are stretched in the first place?

We suggest that the scientific nature of survey research demands that special attention be paid data for secondary use. The overriding philosophy of the NORC Library in its role as conservator of NORC surveys is that the entire history of a survey should be preserved so

that it can be examined by anyone at a later date. The natural history of surveys is required if the claim to science is to be satisfied, especially in that aspect of science which suggests that experiments be repeatable. Of course surveys can never be fully repeated, since time is a variable, but an approximation is possible, and history is needed for that.

DOCUMENTATION AND ITS USES

Ideally, then, the final documentation of a study (not only of data, by the way) would permit replication in order to test the published findings, or, at the very least, to intelligently judge the conclusions drawn from the data. This ideal is not wholly impractical at NORC, since the various stages of surveys do in fact produce written documentation. Thus our codebooks contain information which recognizes the broader nature of survey documentation by including information on the sample, field work and training, question specifications, bibliographies of publications and the usual data location and coding instructions.

Although the view of documentation outlined above does not solve the daily crunch, it does give a basis for deciding what documentation should include. It is too bad that there is really very little pressure for good documentation even from the most sophisticated users. The rare exception is the graduate student or professor working in the area of survey methodology.

In the work of archiving, we are largely dependent on the quality of documentation produced by each survey. Until recently, there were traditions of processing at NORC which made it possible to supplement missing information from what one knew about other studies done at about the same time. With the widespread use of statistical analysis software packages and the transition from operator controlled data processing to operating systems, there has been a loss of information about data processing and data analysis of surveys. The older systems required that records be kept in order to communicate jobs to other persons. The trend now appears to be that the individual researcher or research assistant does most of his or her own work. One cannot, however, expect more documentation than is required by the work organization. The individual researcher requires fewer

records and tends not to produce intelligible information about his work, because there is little need, at that point in the research, to communicate with others. As a consequence the constructed variables and final form of data as used by researchers tend to be unusable. For these reasons the NORC library emphasizes the need to maintain the original file and to document the derivation of important derived variables so that they may be reconstructed using the original data.

Older and more recent surveys pose very different sets of problems for archiving at NORC. An example of each will point out some of these. An example of an older study is the May, 1964 occupational prestige study which is the basis for the Hodge-Siegel-Rossi prestige score (used in the General Social Survey). The most remarkable and revealing thing is that this study was not archived until February, 1976, due to study director reluctance to allow secondary use. In the not so distant future, studies such as these will be virtually irretrievable, for the simple reason that much of the data are saved only on punched cards (which are warped now) with multiple punches, and the collective memories of the people involved with the studies tend to be more and more vague. Almost all NORC data prior to mid-1960s can be expected to contain some multiple punches which require special processing that increases cost (and tediousness, too) of archiving. This problem also makes it difficult to create fresh copies of data. Time, in general, is a crucial factor in data storage, since even tape copies can be expected to become unreadable after a while. There is also the problem of reconstructing documentation for these older studies. In many cases we must rely on memories or personal files to recreate some crucial piece of information. Finally, with the immanent loss of the last few pieces of unit record equipment at NORC, we can expect that the solution for the problems of archiving older data sets will become even more difficult.

The Continuous National Survey (CNS) was archived in December, 1975 and is an example, although extreme, of some problems with more recent studies (this study was conducted from April, 1973 to May, 1974). Although the data are documented and usable, they are only available in the form of SPSS system files or character coded data derived from the SPSS files. The original data were not processed by

the traditional means at NORC. Rather, they were punched in free-field format and processed almost entirely with custom designed software, which was never used again. The original data have been entirely lost; an intermediate form of the character coded data is unusable because of a lack of documentation. There is virtually no way to verify the quality of the SPSS system files (short of going back to the seven or eight thousand questionnaires). A less serious but aggravating problem is the complex documentation required for the system files as opposed to the usually more straight forward material associated with original data.

CONCLUSIONS

Archival problems will continue at NORC -- of that there is no doubt. If there was serious concern with archiving as a part of

the methodology in the social sciences, and perhaps more of a recognition that a scientific enterprise requires good documentation (and the means to get it), we would be more optimistic. As it is, we expect to see more archival problems with recent studies and increasing difficulty with data from older studies.

So that we should not end on such a glum note, let us point out that as a special library mandated with the care of and communication of NORC information and data, we do fairly well, considering the pressures for service with which we must cope. For the future we will continue to be as responsive to our publics as we can be, but we may adjust our priorities and spend more time with the all important job of study documentation so that NORC studies can find their way, at perhaps a faster rate, into the nation's data archives and data libraries.